IN THE CLAIMS

The following is a complete listing of the claims now pending. This listing replaces all earlier versions and listings of the claims.

Claim 1 (currently amended): A document type definition generating method for generating a document type definition of a structured document provided with tags, each tag having an element name for each document element containing document elements of a plurality of document element types, wherein each one of the plurality of document element types has a document element name and each document element has a start tag and an end tag, said method comprising:

a physical structure judging step[[,]] of judging a physical similarity between physical structures of each of the document elements in the structured document the document elements in the structured document, wherein the judging of the physical similarity is based on the physical position of the start tag of each document element in the structured document;

a semantic structure judging step[[,]] of judging a semantic similarity between the document elements semantic structures of each of the tags by comparing a character string form located between the start tag and the end tag of each of the document elements the form of each tagged element; and

a document type definition generating step[[,]] of judging a similarity of the document element tags based on judgment the results [[of]] obtained in said physical structure judging step and said semantic structure judging step, and generating the document type definition which unifies unifying the document element names of [[the]] similar document elements [[tags]].

2

Claim 2 (currently amended): A document type definition generating method according to claim 1, wherein said physical structure judging step ~~comprises~~ includes judging the ~~physical structure~~ physical similarity of the document elements based on an indentation or a blank line in the structured document.

Claim 3 (currently amended): A document type definition generating method according to claim 2, wherein, when the physical ~~structure of the document element~~ similarity of the document elements is judged based on the indentation in the structured document in said physical structure judging step, the judging is performed by excluding the indentation which represents a quotation.

Claim 4 (currently amended): A document type definition generating method according to claim 2, wherein, when the physical ~~structure of the document element~~ similarity of the document elements is judged based on the blank line in the structured document in said physical structure judging step, the judging is performed by excluding ~~the blank line~~ a determined number of blank lines from the structured document ~~in which a description is made~~ wherein the number of blank lines is determined by constantly ~~placing every predetermined number of~~ skipping one or more blank lines.

Claim 5 (canceled)

Claim 6 (currently amended): A document type definition generating method according to claim 1, wherein said semantic structure judging step ~~comprises referring to~~ includes accessing a semantic information database to judge the semantic ~~structure~~

3

similarity of the document element based on a connection of words and phrases in the structured document and word types.

Claim 7 (currently amended): A document type definition generating method according to claim 1, wherein said semantic structure judging step ~~comprises~~ includes judging the semantic ~~structure~~ similarity of the document element based on a meaning represented by [[the]] document element tags surrounding the document element.

Claim 8 (currently amended): A document type definition generating method according to claim 1, wherein said document type definition generating step ~~comprises~~ includes a redundancy removing step of, when the physical structure and the semantic structure of a plurality of document elements having tags different in element name are judged similar in said physical structure judging step and said semantic structure judging step, regarding the document elements as being of the same document element type and excluding one document element name from a document type definition generating object based on the judgment results [[of]] obtained in said physical structure judging step and said semantic structure judging step.

Claim 9 (currently amended): A document type definition generating method according to claim 8, wherein said redundancy removing step ~~comprises~~ includes obtaining similarity degrees concerning agreement degrees of the physical structure and the semantic structure between [[the]] document elements having tags different in document element name, and regarding the document elements as being of the same type when a general

4

similarity value calculated from the similarity degrees <u>in said redundancy removing step</u> is

equal to or [[more]] <u>greater</u> than a predetermined threshold value.

Claim 10 (currently amended): A document type definition generating

method according to claim 1, wherein said document type definition generating step

~~comprises~~ <u>includes</u> a title changing step of, when the physical structure and the semantic

structure of a plurality of document elements having <u>document element</u> tags with the same

<u>document</u> element name are <u>judged to be</u> different <u>in said physical structure judging step and</u>

<u>said semantic structure judging step</u>, regarding the document elements as being of different

<u>document element</u> types and changing one <u>document</u> element name based on the judgment

results [[of]] <u>obtained in</u> said physical structure judging step and said semantic structure

judging step.

Claim 11 (canceled)

Claim 12 (currently amended): A document type definition generating

apparatus for generating a document type definition of a structured document ~~provided with~~

~~tags, each tag having an element name for each document element~~ <u>containing document</u>

<u>elements of a plurality of document element types, wherein each one of the plurality of</u>

<u>document element types has a document element name and each document element has a start</u>

<u>tag and an end tag</u>, said apparatus comprising:

physical structure judging means for judging a <u>physical</u> similarity

between ~~physical structures of each of the document elements in the structured document~~ <u>the</u>

<u>document elements in the structured document, wherein the judging of the physical similarity</u>

is based on the physical position of the start tag of each document element in the structured document;

semantic structure judging means for judging a semantic similarity between the document elements ~~semantic structures of each of the tags~~ by comparing a character string form located between the start tag and the end tag of each of the document elements ~~the form of each tagged element~~; and

document type definition generating means for judging a similarity of the document element tags based on ~~judgment~~ the results of said physical structure judging means and said semantic structure judging means, and generating the document type definition ~~which unifies~~ unifying the document element names of [[the]] similar document elements [[tags]].

Claim 13 (currently amended): A document type definition generating apparatus according to claim 12, wherein said physical structure judging means judges the physical ~~structure~~ similarity of the document ~~element~~ elements based on an indentation or a blank line in the structured document.

Claim 14 (currently amended): A document type definition generating apparatus according to claim 13, wherein said physical structure judging means judges the physical ~~structure~~ similarity of the document ~~element~~ elements based on the indentation by excluding the indentation which represents a quotation.

Claim 15 (currently amended): A document type definition generating apparatus according to claim 13, wherein said physical structure judging means judges the

physical ~~structure~~ similarity of the document ~~element~~ elements based on the blank lines by

excluding [[the]] a determined number of blank lines from [[a]] the structured document ~~in~~

~~which description is made~~ wherein the number of blank lines is determined by constantly

~~placing every predetermined number of~~ skipping one or more blank lines.

Claim 16 (canceled)

Claim 17 (currently amended): A document type definition generating

apparatus according to claim 12, wherein said semantic structure judging means ~~refers to~~

accesses a semantic information database to judge the semantic ~~structure~~ similarity of the

document element based on a connection of words and phrases in the structured document

and word types.

Claim 18 (currently amended): A document type definition generating

apparatus according to claim 12, wherein said semantic structure judging means judges the

semantic ~~structure~~ similarity of the document element based on a meaning represented by

[[the]] document element tags surrounding the document element.

Claim 19 (currently amended): A document type definition generating

apparatus according to claim 12, wherein said document type definition generating means

~~comprises~~ includes redundancy removing means for, when the physical structure and the

semantic structure of a plurality of document elements having tags different in element name

are judged similar by said physical structure judging means and said semantic structure

judging means, regarding the document elements as being of the same document element type

and excluding one <u>document</u> element name from a document type definition generating object

based on the judgment results of said physical structure judging means and said semantic

structure judging means.

Claim 20 (currently amended): A document type definition generating

apparatus according to claim 19, wherein said redundancy removing means obtains similarity

degrees concerning agreement degrees of the physical structure and the semantic structure

between [[the]] document elements having tags different in element name, and regards the

document elements as being of the same <u>document element</u> type when a general similarity

value calculated from the similarity degrees <u>by said redundancy removing means</u> is equal to or

[[more]] <u>greater</u> than a predetermined threshold value.

Claim 21 (currently amended): A document type definition generating

apparatus according to claim 12, wherein said document type definition generating means

~~comprises~~ <u>includes</u> title changing means for, when the physical structure and the semantic

structure of a plurality of document elements having <u>document element</u> tags with the same

element name are <u>judged to be</u> different <u>by said physical structure judging means and said</u>

<u>semantic structure judging means</u>, regarding the document elements as being of different

<u>document element</u> types and changing one <u>document</u> element name based on the judgment

results of said physical structure judging means and said semantic structure judging means.

Claim 22 (canceled)

Claim 23 (currently amended): A computer-readable storage medium

storing a program for controlling a computer to execute a document type definition generation

method for generating document type definition of a structured document ~~provided with tags,~~

~~each tag having an element name for each document element~~ containing document elements

of a plurality of document element types, wherein each one of the plurality of document

element types has a document element name and each document element has a start tag and an

end tag, said program comprising:

code for a physical structure judging step[[,]] of judging a physical

similarity between ~~physical structures of each of the document elements in the structured~~

~~document~~ the document elements in the structured document, wherein the judging of the

physical similarity is based on the physical position of the start tag of each document element

in the structured document;

code for a semantic structure judging step[[,]] of judging a semantic

similarity between the document elements ~~semantic structure of said each document element~~

~~structures of each of the tags~~ by comparing a character string form located between the start

tag and the end tag of each of the document elements ~~the form of each tagged element~~; and

code for a document type definition generating step[[,]] of judging a

similarity of the document element tags based on ~~judgment~~ the results [[of]] obtained by said

physical structure judging [[step]] code and said semantic structure judging [[step]] code, and

generating the document type definition ~~which unifies~~ unifying the document element names

of [[the]] similar document elements [[tags]].

Claim 24 (new): A processing method for processing a structured document

containing document elements of a plurality of document element types, wherein each one of

9

the plurality of document element types has a document element name and each document element has a start tag and an end tag, said method comprising:

an input step of inputting the structured document; and

a judging step of judging the similarity between the document elements by comparing a character string form located between the start tag and the end tag of each document element.

Claim 25 (new): A processing method for processing a structured document containing document elements of a plurality of document element types, wherein each one of the plurality of document element types has a document element name and each document element has a start tag and an end tag, said method comprising:

an input step of inputting the structured document; and

a judging step of judging the similarity between the document elements based on a physical similarity of the document elements according to positions of each document element start tag in the structured document.

Claim 26 (new): A processing apparatus for processing the similarity of document elements in a structured document containing document elements of a plurality of document element types, wherein each one of the plurality of document element types has a document element name and each document element has a start tag and an end tag, said apparatus comprising:

an input device for inputting the structured document; and

a judging device for judging the similarity between the document elements by comparing a character string form located between the start tag and the end tag of each document element.

Claim 27 (new): A processing apparatus for processing the similarity of document elements in a structured document containing document elements of a plurality of document element types, wherein each one of the plurality of document element types has a document element name and each document element has a start tag and an end tag, said apparatus comprising:

an input device for inputting the structured document; and

a judging device for judging the similarity between the document elements based on a physical similarity of the document elements according to positions of each document element start tag in the structured document.